Living with Uncertainty: Toward the Ongoing Normative Assessment of Nanotechnology

Jean-Pierre Dupuy
Alexei Grinbaum
Ecole Polytechnique, France

Nanotechnology's metaphysical research program

It is often asserted that the starting point of nanotechnology was the classic talk given by Feynman (1959), in which he said: "The principles of physics, as far as I can see, do not speak against the possibility of maneuvering things atom by atom...It would be, in principle, possible (I think) for a physicist to synthesize any chemical substance that the chemist writes down. Give the orders and the physicist synthesizes it. How? Put the atoms down where the chemist says, and so you make the substance." Today's champions of nanotech add: "We need to apply at the molecular scale the concept that has demonstrated its effectiveness at the macroscopic scale: making parts go where we want by *putting* them where we want!" (Merke 2003)

This cannot be the whole story. If the essence of nanotechnology were that it manipulates matter on the atomic scale, no new philosophical attitude different from the one to other scientific disciplines would be necessary. Indeed, chemistry has been manipulating matter on the atomic scale for at least the past two centuries. We believe there is indeed some kind of unity behind the nanotech enterprise and the NBIC convergence (Roco et al. 2002); but that this unity lies at the level of the 'metaphysical research program' that underpins such convergence. It is at this level that nanoethics must address novel issues.

Let us recall that Karl Popper, following the lead of Emile Meyerson (1927), defined the notion of metaphysical research program as a set of ideas and worldviews that underlie any particular scientific research agenda. The positivist philosophy that drives most of modern science (and much of contemporary philosophy) takes 'metaphysics' to be a meaningless quest for answers to unanswerable questions. However, Popper showed that there is no scientific (or, for that matter, technological) research program that would not rest on a set of general presuppositions about the structure of the world. To be sure, those metaphysical views are not empirically testable and they are not amenable to 'falsification.' However, this does not imply that they are not of less importance or that they do not play a fundamental role in the advancement of science. Those who deny metaphysics simply render it invisible, and it is very likely that their hidden metaphysics is bad or inconsistent. To the amazement of those who mistook him for a positivist, Karl Popper claimed that the philosopher or historian of science's task was twofold: first, unearth and make visible the metaphysical ideas that lie underneath scientific programs in order to make them amenable to criticism; second, to proceed to a critical examination of those metaphysical theories, in a way that is different from the criticism of scientific theories, since no empirical testing is here possible, but nevertheless rational.

Our claim is that the major ethical issues raised by the nanotech enterprise and the NBIC convergence are novel and that they originate in the metaphysical research program on which such convergence rests. In order to substantiate this claim, we submit that the origin of the NBIC convergence is to be sought in another classic conference, the one John von Neumann gave at Caltech (1948) on complexity and self-reproducing automata.

Turing's and Church's theses were very influential at the time, and they had been supplemented by cyberneticians Warren McCulloch and Walter Pitts' major finding on the properties of neural networks (Dupuy 2000b, pp. 68-69). Cybernetics' credo was then: every behavior that is unambiguously describable in a finite number of words is computable by a network of formal neurons—a remarkable statement, as von Neumann recognized. However, he put forward the following objection: is it reasonable to assume as a practical matter that our most complex behaviors are describable in their totality, without ambiguity, using a finite number of words? In specific cases it is always possible: our capacity, for example, to recognize the same triangular form in two empirical triangles displaying differences in line, size, and position can be so described. But would this be possible if it were a matter of globally characterizing our capacity for establishing 'visual analogies'? In that case, von Neumann conjectured, it may be that the simplest way to describe a behavior is to describe the structure that generates it. It is meaningless, under these circumstances, to 'discover' that such a behavior can be embodied in a neural network since it is not possible to define the behavior other than by describing the network itself. To take an illustration:

> The unpredictable behaviour of nanoscale objects means that engineers will not know how to make nanomachines until they actually start building them (*The Economist*, March 2003).

Von Neumann thus posed the question of complexity, foreseeing that it would become the great question for science in the future. Complexity implied for him, in this case, the futility of the constructive approach of McCulloch and Pitts, which reduced a function to a structure, thus leaving unanswered the question of what a complex structure is capable.

It was in the course of his work on automata theory that von Neumann was to refine this notion of complexity. Assuming a magnitude of a thermodynamic type, he conjectured that below a certain threshold it would be degenerative, meaning that the degree of organization could only decrease, but that above this threshold an increase in complexity became possible. Now this threshold of complexity, he supposed, is also the point at which the structure of an object becomes simpler than the description of its properties. Soon, von Neumann prophesied, the builder of automata would find himself as helpless before his creation as we feel ourselves to be in the presence of complex natural phenomena (Dupuy 2000b).

At any rate, von Neumann was thus founding the so-called *bottom-up approach*. In keeping with that philosophy, the engineers will not be any more the ones who devise and design a structure capable of fulfilling a function that has been assigned to them. The engineers of the future will be the ones who know they are successful when they are surprised by their own creations. If one of your goals is to reproduce life, to fabricate life, you have to be able to simulate one of its most essential properties, namely the capacity to complexify.

Admittedly, not all of nanotech falls under the category of complexity. Most of today's realizations are in the field of nanomaterials and the problems they pose have to do with toxicity. However, as a recent report by the European Commission says, "the powerful heuristic of Converging Technologies will prove productive even if it is or should be realized to a small extent only" (Nordmann 2004). The effects that pose ethical problems are not only the effects of technology *per se*, but also the effects of the metaphysical ideas that drive technology, whether technological realizations see the light of day or not. We are here mainly interested in these. Among them the novel kind of uncertainty associated with an ambition or a dream to set off complex phenomena looms large.

Towards a novel concept of prudence

In her masterly study of the frailties of human action, Hannah Arendt brought out the fundamental paradox of our time: as human powers increase through technological progress, we are less and less equipped to control the consequences of our actions. From the start, a long excerpt is worth quoting, as its relevance for our topic cannot be overstated—and we should keep in mind that this was written in 1958:

> ...the attempt to eliminate action because of its uncertainty and to save human affairs from their frailty by dealing with them as though they were or could become the planned products of human making has first of all resulted in channeling the human capacity for action, for beginning new and spontaneous processes which without men never would come into existence, into an attitude toward nature which up to the latest stage of the modern age had been one of exploring natural laws and fabricating objects out of natural material. To what extent we have begun to *act into nature*, in the literal sense of the word, is perhaps best illustrated by a recent casual remark of a scientist who quite seriously suggested that "*basic research is when I am doing what I don't know what I am doing.*"
>
> This started harmlessly enough with the experiment in which men were no longer content to observe, to register, and contemplate whatever nature was willing to yield in her own appearance, but began to prescribe conditions and to provoke natural processes.

> What then developed into an ever-increasing skill in *unchaining elemental processes*, which, without the interference of men, would have lain dormant and perhaps never have come to pass, has finally ended in a veritable art of *'making' nature*, that is, of creating 'natural' processes which without men would never exist and which earthly nature by herself seems incapable of accomplishing...
>
> The very fact that natural sciences have become exclusively sciences of process and, in their last stage, *sciences of potentially irreversible, irremediable 'processes of no return'* is a clear indication that, whatever the brain power necessary to start them, *the actual underlying human capacity which alone could bring about this development is no 'theoretical' capacity, neither contemplation nor reason, but the human ability to act*—to start new unprecedented processes whose outcome remains uncertain and unpredictable whether they are let loose in the human or the natural realm.
>
> In this aspect of action...processes are started whose outcome is unpredictable, so that *uncertainty rather than frailty becomes the decisive character of human affairs* (Arendt 1958, 230-232; our emphasis).

No doubt that with an incredible prescience this analysis applies perfectly well to the NBIC convergence, in particular on two scores. Firstly, the ambition to (re-)make nature is an important dimension of the metaphysical underpinnings of the field. If the NBIC converging technologies purport to take over Nature's and Life's job and become the engineers of evolution, it is because they have redefined Nature and Life in terms that belong to the realm of artifacts. See how one of their most vocal champions, Damien Broderick, rewrites the history of life, or, as he puts it, of "living replicators":

> Genetic algorithms *in planetary numbers lurched about on the surface of the earth and under the sea, and indeed as we now know deep within it, for billions of years, replicating and mutating and being winnowed via the success of their expressions—that is, the bodies they* manufactured*, competing for survival in the macro world. At last, the entire living ecology of the planet has* accumulated, and represents a colossal quantity of compressed, schematic information (2001, p. 116).

Once life has thus been transmogrified into an artifact, the next step is to ask oneself whether the human mind couldn't do better. The same author asks rhetorically, "Is it likely that nanosystems, designed by human minds, will bypass all this Darwinian wandering, and leap straight to design success?" (p. 118)

Secondly, as predicted by von Neumann, it will be an inevitable temptation, not to say a task or a duty, for the nanotechnologists of the future to set off processes upon which they have no control. The sorcerer's apprentice myth must be updated: it is neither by error nor by terror that Man will be dispossessed of his own creations but by design.

There is no need for Drexlerian self-assemblers to come into existence for this to happen. The paradigm of *complex, self-organizing systems* envisioned by von Neumann is stepping ahead at an accelerated pace, both in science and in technology. It is in the process of shoving away and replacing the old metaphors inherited from the cybernetic paradigm, like the ones that treat the mind or the genome as computer programs. In science, the central dogmas of molecular biology received a severe blow on two occasions recently. First, with the discovery that the genome of an adult, differentiated cell can be 'reprogrammed' with the cooperation of maternal cytoplasm—hence the technologies of nucleus transfer, including therapeutic and reproductive cloning. Secondly, with the discovery of prions, which showed that self-replication does not require DNA. As a result, the sequencing of the human genome appears to be not the end of the road but its timid beginning. Proteinomics and complexity are becoming the catchwords in biology, relegating genomics to the realm of passé ideas.

In technology, new feats are being flaunted every passing week. Again, the time has not come—and may never come—when we manufacture self-replicating machinery that mimics the self-replication of living materials. However, we are taking more and more control of living materials and their capacity for self-organization and we use them to perform mechanical functions.

Examples are plenty. To give just one: In November 2003, scientists in Israel built transistors out of carbon nanotubes using DNA as a template. A Technion-Israel scientist said, "What we've done is to bring biology to *self-assemble an electronic device* in a test tube...The DNA serves as a scaffold, a template that will determine where the carbon nanotubes will sit. That's the beauty of using biology" (Chang 2003).

From a philosophical point of view the key issue is to develop new concepts of prudence that are suited to this novel situation. A long time ago Aristotle's *phronesis* was dislodged from its prominent place and replaced with the modern tools of the probability calculus, decision theory, the theory of expected utility, etc. More qualitative methods, such as futures studies, 'Prospective', and the scenario method were then developed to assist decision-making. More recently, the precautionary principle emerged on the international scene with an ambition to rule those cases in which uncertainty is mainly due to the insufficient state of our scientific knowledge. We believe that none of these tools is appropriate for tackling the situation that we are facing now.

From the outset we make it explicit that our approach is inherently normative. German philosopher Hans Jonas cogently explains why we need a radically new ethics to rule our relation to the future in the "technological age" (Jonas 1985). This "Ethics of the Future" (*Ethik* für *die Zukunft)*—meaning not a future ethics, but an ethics *for* the future, for the sake of the future, i.e. the future must become the major object of our concern—starts from a philosophical aporia. Given the magnitude of the possible consequences of our technological choices, it is an absolute obligation for us to try and anticipate those consequences, assess them, and ground our choices on this assessment. Couched in philosophical parlance, this is tantamount to saying that when the stakes are high, as in predicting the future, none of the normative ethics that are available is up to the challenge. Virtue ethics is manifestly insufficient since the problems ahead have very little to do with the fact that scientists or engineers are beyond moral reproach or not. Deontological doctrines do not fare much better since they evaluate the rightness of an action in terms of its conformity to a norm or a rule, for example to the Kantian categorical imperative: we are now well acquainted with the possibility that 'good' (e.g. democratic) procedures lead one into an abyss. As for consequentialism—i.e. the set of doctrines that evaluate an action based on its consequences for all agents concerned—it treats uncertainty as does the theory of expected utility, namely by ascribing probabilities to uncertain outcomes. Hans Jonas argues that doing so has become morally irresponsible. The stakes are so high that we must set our eyes on the worst-case scenario and see to it that it never sees the light of day.

However, the very same reasons that make our obligation to anticipate the future compelling, make it impossible for us to do so. Unleashing complex processes is a very perilous activity that both demands certain foreknowledge and prohibits it. Indeed, one of the very few unassailable ethical principles is that *ought* implies *can*. There is no obligation to do that which one cannot do. However, we do have here an ardent *obligation* that we can*not* fulfil: anticipating the future. We cannot but violate one of the foundations of ethics.

What is needed is a novel approach to the future, neither scenario nor forecast. We submit that what we call *ongoing normative assessment* is a step in that direction. In order to introduce this new concept we need to take a long detour into the classic approaches to the problems raised by uncertainty.

Uncertainty Revisited

Shortcomings of the Precautionary Principle

The precautionary principle triumphantly entered the arena of methods to ensure prudence. All the fears of our age seem to have found shelter in the word 'precaution'. Yet, in fact, the conceptual underpinnings of the notion of precaution are extremely fragile.

Let us recall the definition of the precautionary principle formulated in the French Barnier law: "The absence of certainties, given the current state of scientific and technological knowledge, must not delay the adoption of effective and proportionate preventive measures aimed at forestalling a risk of grave and irreversible damage to the environment at an economically acceptable cost" (1995). This text is torn between the logic of economic calculation and the awareness that the context of decision-making has radically changed. On one side, the familiar and reassuring notions of effectiveness, commensurability and reasonable cost; on the other, the emphasis on the uncertain state of knowledge and the gravity and irreversibility of damage. It would be all too easy to point out that if uncertainty prevails, no one can say what would be a measure proportionate (by what coefficient?) to a damage that is unknown, and of which one therefore cannot say if it will be grave or irreversible; nor can anyone evaluate what adequate prevention would cost; nor say, supposing that this cost turns out to be 'unacceptable,' how one should go about choosing between the health of the economy and the prevention of the catastrophe.

One serious deficiency, which hamstrings the notion of precaution, is that it does not properly gauge the type of uncertainty with which we are confronted at present. The report on the precautionary principle prepared for the French Prime Minister (Kourilsky & Viney 2000) introduces what initially appears to be an interesting distinction between two types of risks: 'known' risks and 'potential' risks. It is on this distinction that the difference between prevention and precaution is said to rest: precaution would be to potential risks what prevention is to known risks. A closer look at the report in question reveals 1) that the expression 'potential risk' is poorly chosen, and that what it designates is not a risk waiting to be realized, but a hypothetical risk, one that is only a matter of conjecture; 2) that the distinction between known risks and, call them this way, hypothetical risks corresponds to an old standby of economic thought, the distinction that John Maynard Keynes and Frank Knight independently proposed in 1921 between risk and uncertainty. A risk can in principle be quantified in terms of objective probabilities based on observable frequencies; when such quantification is not possible, one enters the realm of uncertainty.

The problem is that economic thought and decision theory underlying it were destined to abandon the distinction between risk and uncertainty as of the 1950s in the wake of the exploit successfully performed by Leonard Savage with the introduction of the concept of subjective probability and the corresponding philosophy of choice under conditions of uncertainty: Bayesianism. In Savage's approach, probabilities no longer correspond to any sort of objective regularity present in nature, but simply to the coherent sequence of a given agent's choices. In philosophical language, every uncertainty is treated as *epistemic* uncertainty, meaning an uncertainty associated with the agent's state of knowledge. It is easy to see that introduction of subjective probabilities erases Knight's distinction between uncertainty and risk, between risk and the risk of risk, between precaution and

prevention. If a probability is unknown, all that happens is that a probability distribution is assigned to it subjectively. Then further probabilities are calculated following the Bayes rule. No difference remains compared to the case where objective probabilities are available from the outset. Uncertainty owing to lack of knowledge is brought down to the same plane as intrinsic uncertainty due to the random nature of the event under consideration. A risk economist and an insurance theorist do not see and cannot see any essential difference between prevention and precaution and, indeed, reduce the latter to the former. In truth, one observes that applications of the 'precautionary principle' generally boil down to little more than a glorified version of 'cost-benefit' analysis.

Our situation with respect to new threats is different from the above-discussed context. The novel feature this time is that although uncertainty is objective, we are not dealing with a random occurrence either. This is because each of the future great discoveries or of the future catastrophes must be treated as a *singular event*. Neither random, nor uncertain in the usual epistemic sense, the type of 'future risk' that we are confronting is a monster from the standpoint of classic distinctions. Indeed, it merits a special treatment, which the precautionary principle is incapable of giving.

When the precautionary principle states that the "absence of certainties, given the current state of scientific and technical knowledge, must not delay etc.," it is clear that it places itself from the outset within the framework of epistemic uncertainty. The assumption is that we know we are in a situation of uncertainty. It is an axiom of epistemic logic that if I do not know P, then I know that I do not know P. Yet, as soon as we depart from this framework, we must entertain the possibility that we do not know that we do not know something. In cases where uncertainty is such that it entails that uncertainty itself is uncertain, it is impossible to know whether or not the conditions for application of the precautionary principle have been met. If we apply the principle to itself, it will invalidate itself before our eyes.

Moreover, "given the current state of scientific and technical knowledge" implies that a scientific research effort could overcome the uncertainty in question, whose existence is viewed as purely contingent. It is a safe bet that a 'precautionary policy' will inevitably include the edict that research efforts must be pursued—as if the gap between what is known and what needs to be known could be filled by a supplementary effort on the part of the knowing subject. But it is not uncommon to encounter cases in which the progress of knowledge comports an increase in uncertainty for the decision-maker, a thing inconceivable within the framework of epistemic uncertainty. Sometimes, to learn more is to discover hidden complexities that make us realize that the mastery we thought we had over phenomena was in part illusory.

Society is a participant

From the point of view of mathematics of complex systems one can distinguish several different sources of uncertainty. Some of them appear in almost any analysis of uncertainties; others are taken into account quite rarely.

Presence of tipping points, i.e. such points on the system's landscape of trajectories that trigger an abrupt fall of the system into states completely different from the states that the system had previously occupied, is one of the reasons why uncertainty is not amenable to the concept of probability. As long as the system remains far from the threshold of the catastrophe, it may be handled with impunity. Here cost-benefit analysis of risks is bound to produce a banal result, because the trajectory is predictable and no surprises can be expected. To give an example, this is the reason why humanity was able to blithely ignore, for centuries, the impact of its mode of development on the environment. But as the critical thresholds grow near, cost-benefit analysis, previously a banality, becomes meaningless. At that point it is imperative not to enter the area of critical change at any cost, if one, of course, wants to avoid the crisis and sustain the smooth development. We see that for reasons having to do, not with a temporary insufficiency of our knowledge, but with the structural properties of complex systems, economic calculation is of little help.

We now turn to another source of uncertainty that appears in the case of systems in whose development participates the human society. Technology here is just one example. To these systems the usual techniques for anticipating the future, as discussed in the next section, are inapplicable. The difficulty comes from the fact that, in general, any system where the society plays an active role is characterized by the impossibility to dissociate the observed part of the system ('the sphere of technology') from the observer ('society at large'), who himself is influenced by the system and must be viewed as one of its components. In a usual setting, the observer looks at the system that he studies from an external point, and both the observer and the system evolve in linear physical time. The observer can then treat the system as independent from the act of observation and can create scenarios in which this system will evolve in linear time. Not so if the observer can influence the system and, in turn, be influenced by it (Figure 1). What evolves as a whole in linear time is now a conglomerate, a composite system consisting of both the complex system and the observer. However, the evolution of the composite system in the linear time becomes of no interest for us, for the act of observation is performed by the observer who is a part of the composite system; the observer himself is now *inside the big whole*, and his point of view is no more an external one. The essential difference is that the observer and the complex system enter into a network of complex relations with each other, due to mutual influence. In science such composite systems are referred to as self-referential systems. They were first studied by von Neumann in his famous book on the theory of self-reproducing automata, which consequently gave rise to a whole new direction of mathematical research.

According to Breuer's theorem, the observer involved in a self-referential system can never have full information on the state of the system. This is a fundamental source of uncertainty in the analysis of complex systems that involve human action. We should take very seriously the idea that there is a "co-evolution of technology and society" (Rip *et al*. 1995). The dynamics of technological development is embedded in society. The consequences of the development of nanotechnology will concern society as well as technology itself. Technology and society shape one another. One can then prove mathematically that the society cannot know with certainty where the technological progress will take it nor make any certain predictions about its own future state.

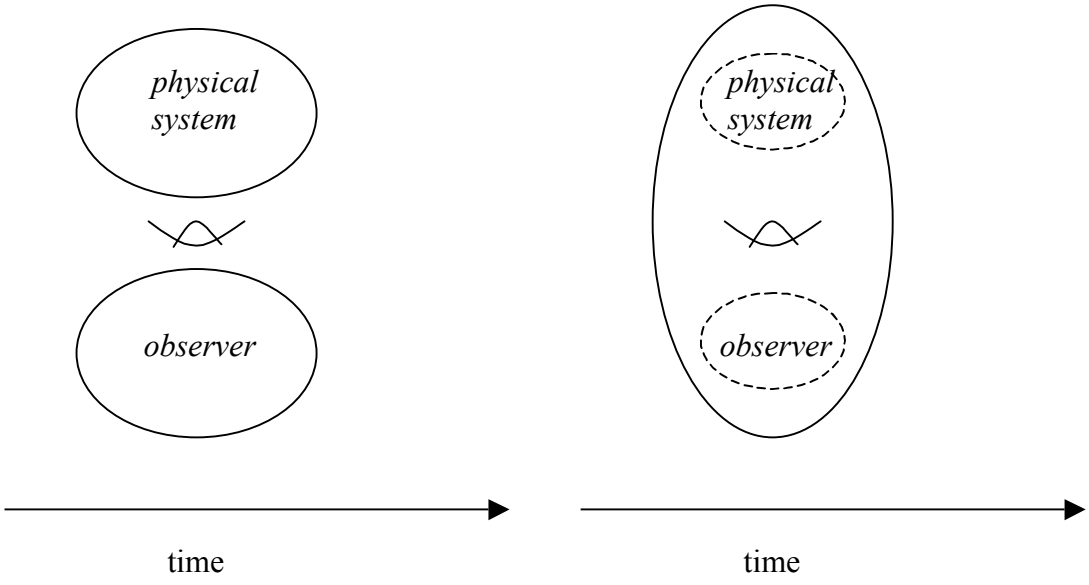

Figure 1. An external observer and an observer-participant.

Projected time

It is a gross simplification to treat the sphere of technology as if it developed only according to its internal logic. Political decision-making and the opinion of the society influence research. The decisions that will be made or not, such as various moratoria and bans, will have a major impact on the evolution of research. Scientific ethics committees would have no *raison d'être* otherwise. If many scientists and experts ponder over the strategic and philosophical questions, it is not only out of curiosity; rather, it is because they wish to exert an influence on the actions that will be taken by the politicians and, beyond, the peoples themselves.

These observations may sound trivial. It is all the more striking that they are not taken into account, most of the time, when it comes to anticipating the evolution of research. When they are, it is in the manner of control theory: human decision is treated as a parameter, an independent or exogenous variable, and not as an endogenous variable. Then, a crucial causal link is missing: the motivational link. It is obvious that human decisions that will be made will depend, at least in part, on the kind of anticipation of the future of the system, this anticipation being made public. And this future will depend, in turn, on the decisions that will be made. A causal loop appears here, that prohibits us from treating human action as an independent variable. Thus, research and technology are systems in which society is a participant.

By and large there are three ways of anticipating the future of a human system, whether purely social or a hybrid of society and the physical world. The first one we call *Forecasting*. It treats the system as if it were a purely physical system. This method is legitimate whenever it is obvious that anticipating the future of the system has no effect whatsoever on the future of the system.

The second method we call, in French, '*Prospective'*. Its most common form is the scenario method. Ever since its beginnings the scenario approach has gone to great lengths to distinguish itself from mere forecast or foresight, held to be an extension into the future of trends observed in the past. We can forecast the future state of a physical system, it is said, but not what we shall decide to do. It all started in the 1950s when a Frenchman, Gaston Berger, coined the term '*Prospective'*—a substantive formed in analogy with 'Retrospective'—to designate a new way to relate to the future. That this new way had nothing to do with the project or the ambition of anticipating, that is, *knowing* the future, was clearly expressed in the following excerpt from a lecture given by French philosopher Bertrand de Jouvenel. In "Of Prospective" he said:

> It is unscholarly perforce because there are no facts on the future. Cicero quite rightly contrasted past occurrences and occurrences to come with the contrasted expressions *facta* and *futura*: *facta*, what is accomplished and can be taken as solid; *futura*, what shall come into being, and is as yet 'undone,' or fluid. This contrast leads me to assert vigorously: '*there can be no science of the future*.' *The future is not the realm of the 'true or false' but the realm of 'possibles.' (de Jouvenel 1964)*

Another term coined by Jouvenel that was promised to a bright future was '*Futuribles*,' meaning precisely the open diversity of *possible futures*. The exploration of that diversity was to become the scenario approach.

A confusion spoils much of what is being offered as the justification of the scenario approach. On the one hand, the alleged irreducible multiplicity of the 'futuribles' is explained as above by the *ontological indeterminacy* of the future: since we 'build,'

'invent' the future, there is nothing to know about it. On the other hand, the same multiplicity is interpreted as the inevitable reflection of our inability to know the future *with certainty*. The confusion of ontological indeterminacy with epistemic uncertainty is a very serious one. From what we read in the literature on nanotechnology, we got the clear impression that the emphasis is put on epistemic uncertainty, but only up to the point where human action is introduced: then the scenario method is used to explore the sensitivity of technological development to human action.

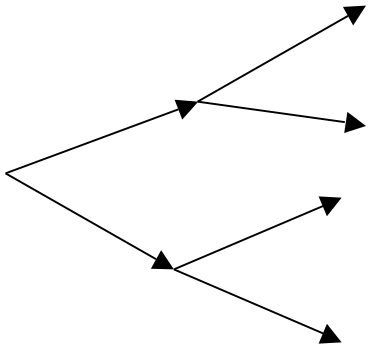

Figure 2. Occurring time

The temporality that corresponds to Prospective or the scenario approach is the familiar decision tree. We call it *occurring time* (Figure 2). It embodies the familiar notions that the future is open and the past is fixed. In short, time in this model is the usual linear one-directional time arrow. It immediately comes to mind that, as we have stated above, linear time does not lead to the correct type of observation and prediction if the observer is an *observer-participant*. This is precisely the case with the society at large and its technology, and, consequently, one must not expect a successful predictive theory of the latter to operate in the linear occurring time.

We submit that occurring time is not the only temporal structure we are familiar with. Another temporal experience is ours on a daily basis. It is facilitated, encouraged, organized, not to say imposed by numerous features of our social institutions. All around us, more or less authoritative voices are heard that proclaim what the more or less near future will be: the next day's traffic on the freeway, the result of the upcoming elections, the rates of inflation and growth for the coming year, the changing levels of greenhouse gases, etc. The *futurists* and sundry other prognosticators know full well, as do we, that this future they announce to us as if it were written in the stars is, in fact, a future of our own making. We do not rebel against what could pass for a metaphysical scandal (except, on occasion, in the voting booth). It is the coherence of this mode of coordination with regard to the future that we have endeavored to bring out, under the name of projected time (Figure 3).

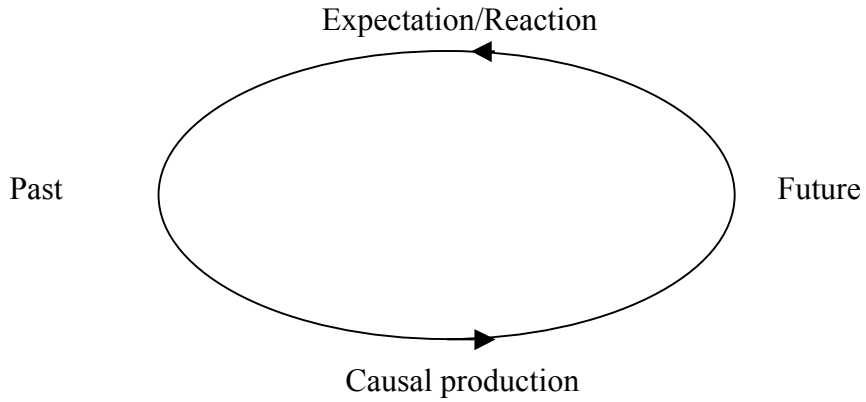

Figure 3. Projected time

To return to the three ways of anticipating the future, the foresight method can be said to be a view of an independent observer from outside the physical system. Counter-argument to it is that in reality the observer is not independent and has a capacity to act as to produce causal effects on the system. The second way of anticipation, 'Prospective,' or its version such as the scenario approach, is a view on the system where the observer is not independent any more, but the view itself is still taken from outside the system. Thus, the one who analyzes and predicts is the same agent as the one who acts causally on the system. As explained in the previous section, this fact entails a fundamental limit on the capacities of the anticipator. What is needed, therefore, is a replacement of the linear occurring time with a different point of view. This means taking seriously the fact that the system involves human action and requiring that predictive theory accounts for this. It is only such a theory that will be capable of providing a sound ground for non-self-contradictory, coherent anticipation. A *sine qua non* must be respected for that coherence to be the case: a *closure condition*, as shown on the graph. Projected time takes the form of a loop, in which past and future reciprocally determine each other. It appears that the metaphysics of projected time differs radically from the one that underlies occurring time, as counterfactual relations run counter causal ones: the future is fixed and the past depends counterfactually upon the future.

To foretell the future in projected time, it is necessary to seek the loop's *fixed point*, where an expectation (on the part of the past with regard to the future) and a causal production (of the future by the past) coincide. The predictor, *knowing that his prediction is going to produce causal effects in the world*, must take account of this fact if he wants the future to confirm what he foretold. Therefore the point of view of the predictor has more to it than a view of the human agent who merely produces causal effects. By contrast, in the scenario

('prospective') approach the self-realizing prophecy aspect of predictive activity is not taken into account.

We will call *prophecy* the determination of the future in projected time, by reference to the logic of self-fulfilling prophecy. Although the term has religious connotations, let us stress that we are speaking of *prophecy* here in a purely secular and technical sense. The prophet is the one who, prosaically, seeks out the fixed point of the problem, the point where *voluntarism achieves the very thing that fatality dictates*. The prophecy includes itself in its own discourse; it sees itself realizing what it announces as destiny. In this sense, as we said before, prophets are legion in our modern democratic societies, founded on science and technology. What is missing is the realization that this way of relating to the future, which is neither building, inventing or creating it, nor abiding by its necessity, requires a special metaphysics, which is precisely provided by what we call projected time (Dupuy 1989; 1992; 1998; 2000a).

Cognitive Barriers

The description of the future determines the future

If the future depends on the way it is anticipated and this anticipation being made public, every determination of the future must take into account the causal consequences of the language that is being used to describe the future and how this language is being received by the general public, how it contributes to shaping public opinion, and how it influences the decision-makers. In other terms, the very description of the future is part and parcel of the determinants of the future. This self-referential loop between two distinct levels, the epistemic and the ontological, is the signature of human affairs. Let us observe that this condition provides us with a criterion for determining which kinds of description are acceptable and which are not: the future *under that description* must be a fixed point of the self-referential loop that characterizes projected time.

Any inquiry on the kind of uncertainty proper to the future states of the co-evolution between technology and society must therefore include a study of the linguistic and cognitive channels through which descriptions of the future are made, transmitted, conveyed, received, and made sense of. This is a huge task, and we will limit ourselves here to two dimensions that seem to us of special relevance for the study of the impact of the new technology: the aversion to not knowing, and the impossibility to believe. A third such dimension that we do not discuss here is the certainty effect studied by Tversky and Kahneman. This effect consists in a practical observation that certainty exaggerates the aversiveness of losses that are certain relative to losses that are merely probable.

Aversion to not knowing

In 1950s, soon after Savage's work, a debate on the subjective probabilities was initiated by Maurice Allais. Allais intended to show that Savage's axioms are very far from what one observes, in economics, in practical decision-making contexts. Soon an example was proposed, a version of which is known under the name of Ellsberg paradox (Ellsberg 1961). The key idea of Allais and, later on, of Ellsberg is that there exists aversion to not knowing. Not knowing must be understood as the opposite of knowing, negation of a certain ascribed property, and must be differentiated from the unknown or ignorance. Ignorance presupposes that something can possibly be known, while here we are concerned with a situation of not knowing and not being able to know, because of the game conditions or because of some real-life factors. Aversion to not knowing can take the form of aversion to uncertainty in situations where uncertainty means epistemic uncertainty according to Frank Knight's distinction between risk and uncertainty. However, as a general principle aversion to not knowing exceeds the conceptual limits of Savage's theory.

The Ellsberg paradox is an example of a situation where agents would irrationally prefer the situation with some information to a situation without any information, although it is rational to prefer to avert from information. Consider two urns, A and B (Figure 4). It is known that in urn A there are exactly ten red balls and ten black balls. About urn B it is only said that it contains twenty balls, some red and some black. A ball from each urn is to be drawn at random. Free of charge, a person can choose one of the two urns and then place a bet on the colour of the ball that is drawn. According to Savage's theory of decision-making, urn B should be chosen even though the fraction of balls is not known. Probabilities can be formed subjectively, and a bet shall be placed on the subjectively most likely ball colour. If subjective probabilities are not fifty-fifty, a bet on urn B will be strictly preferred to one on urn A. If the subjective probabilities are precisely fifty-fifty then the decision-maker will be indifferent. Contrary to the conclusions of Savage's theory, Ellsberg argued that a strict preference for urn A is plausible because the probability of drawing a red or black ball is known in advance. He surveyed the preferences of an elite group of economists to lend support to this position and found that his view was right and that there was evidence against applicability of Savage's axioms. Thus, the Ellsberg paradox challenges the appropriateness of the theory of subjective probability.

We shall also say that the Ellsberg paradox challenges the usual assumption that human decision-makers are probability calculators. Indeed, had one given himself the task of assessing the problem with urns from the point of view of probabilities, it would be inevitable to make use of the Bayes rule and thus conclude that urn B is the preferred choice. But, as shown by Ellsberg, aversion to not knowing is a stronger force than the tendency to calculate probabilities. Aversion to not knowing therefore erects a cognitive barrier that separates human decision-maker from the field of rational choice theory.

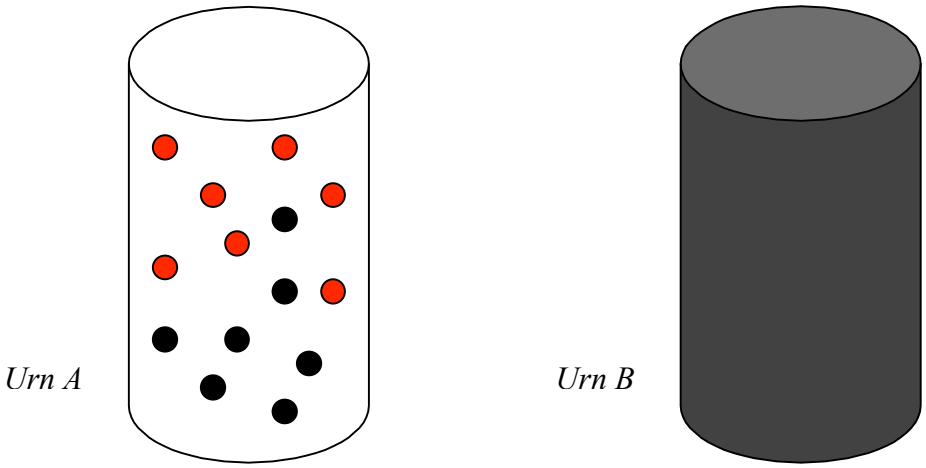

Figure 4. The Ellsberg paradox

Impossibility of believing

Let us again return to the precautionary principle. By placing the emphasis on scientific uncertainty, it misconstrues the nature of the obstacle that keeps us from acting in the face of catastrophe. The obstacle is not just uncertainty, scientific or otherwise; it is equally, if not a more important component, the impossibility of believing that the worst is going to occur. Contrary to many the basic assumption of epistemic logic, one can know that P but still not believe in P.

Pose the simple question as to what the practice of those who govern us was before the idea of precaution arose. Did they institute policies of *prevention,* the kind of prevention with respect to which precaution is supposed to innovate? Not at all. They simply waited for the catastrophe to occur before taking action—as if its coming into existence constituted the sole factual basis on which it could be legitimately foreseen, too late of course. We submit that there exists a deep cognitive basis for such a behaviour, which is exhibited by human decision makers in a situation when they know that a singular event, like a catastrophe, stands right behind the door. In these circumstances arises a cognitive barrier of the impossibility to believe in the catastrophe.

To be sure, there are cases where people do see a catastrophe coming and do adjust. That just means that the cognitive barrier in question is not absolute and can be overcome. We will introduce further a method that makes such overcoming more likely. However, by and large, even when it is known that it is going to take place, a catastrophe is not credible. On the basis of numerous examples, an English researcher David Fleming identified what he called the “inverse principle of risk evaluation”: the propensity of a

community to recognize the existence of a risk seems to be determined by the extent to which it thinks that solutions exist (Fleming 1996). There is no subjective or objective probability calculus here; knowing that P but not believing in P has a different origin.

What could this origin be? Observe first that the aversion to not knowing and the impossibility to believe do not go unconnected. Both are due to the fact that human action as cognitive decision-making process vitally depends on having information. Cognitive agents cannot act without having information that they rely upon, and the experience from which they build analogies with a current situation. Consequently, a fundamental cognitive barrier arises, which is that if an agent does not have information or experience, then he does not take action, a situation that for an outsider appears as paralysis in decision-making. Aversion to not knowing is caused by the cognitive barrier but the agent, like in the Ellsberg paradox, is forced to act. He then chooses an action which is not rational but which escapes to the largest degree the situation of not having information. Were the agent allowed not to act at all, as in real life situations, the most probable outcome becomes the one of paralysis. When the choice is between the relatively bad, the unknown, and doing nothing, the last option happens to be the most attractive one. If it is dropped and the choice is just between the relatively bad and the unknown, relatively bad may turn out to be the winner. To summarize, we argue that a consequence of the cognitive barrier is that if in a situation of absence of information and of the singular character of the coming event there is a possibility not to act, this will be the agent's preference. Standing face to face with a catastrophe or a dramatic change in life, most people become paralyzed. As cognitive agents, they have no information, no experience, and no practical know-how concerning the singular event, and the cognitive barrier precludes the human decision-maker from action.

Another consequence of the cognitive barrier is that if an agent is forced to act, then he will do his best to acquire information. Even though it may later be found out that he had made wrong decisions or his action had not been optimal, in the process of decision-making itself the cognitive barrier dictates that the agent collects as much information as he can get and acts upon it. Reluctance to bring in available information or, yet more graphically, refusal to look for information are by themselves special decisions and require that the agent consciously chooses to tackle the problem of the quality and quantity of information that he wants to act upon. If the agent does so, i.e. if he gives himself the task to analyze the problem of necessary vs. superficial information, then it is comprehensible that the agent would refuse to acquire some information, as does the rational agent in the Ellsberg paradox. But if the meta-analysis of the preconditions of decision-making is not undertaken, then the agent will naturally tend to collect at least some information that is available on the spot. Such is the case in most real life situations. Consequently, the cognitive barrier entails that the directly available information is viewed as relevant to decision-making; if there is no such information, then the first thing-to-do is to look for one.

Cognitive barrier in its clear-cut form applies to situations where one faces a choice between total absence of information and availability of at least some knowledge. The reason why agents have no information on an event and its consequences is usually that this event is a singular event. Singular events, by definition, mean that the agent cannot use his previous experience for analyzing the range of possible outcomes and for evaluating particular outcomes in this range. To enter into Savage's rational decision-making process, agents require previous information or experience that allow them to form priors. If information is absent or is such that no previous experiential data is available, the process is easily paralyzed. Contrary to the prescription of the theory of subjective probabilities, in a situation of absence of information real cognitive agents do not choose to set priors arbitrarily. To them, selecting probabilities and even starting to think probabilistically without any reason to do so appears as purely irrational and untrustworthy. Independently of the projected positive or negative outcome of a future event, if it is a singular event, then cognitive agents stay away from the realm of subjective probabilistic reasoning and are led to paralysis.

Now, our immediate concern becomes to offer a way of functioning, which is capable of bringing the agents back to operational mode from the dead end of cognitive paralysis.

Methodology of ongoing normative assessment

The methodology that we propose is different from a one-time probabilistic analysis that is devoted to constructing a range of scenarios, all developing in the linear time which forks into a multitude of branches, and choosing 'the best', whatever the criterion. Our method does not rest on the application of an *a priori* principle, such as the Precautionary Principle. We submit that no principle can do the job of dealing with the kind of uncertainty that the new technological wave generates. What we propose can be viewed as a practice, rather than a principle, as a *way of life* or a procedural prescription for all kinds of agents: from a particular scientist and a research group to the whole of the informed society, telling them how to proceed with questions regarding the future, on a regular basis in course of their usual work.

Our methodology is a methodology of ongoing normative assessment. *It is a matter of obtaining through research, public deliberation, and all other means, an image of the future sufficiently optimistic to be desirable and sufficiently credible to trigger the actions that will bring about its own realization.* The sheer phrasing of the methodology suggests that it rests on the metaphysics of projected time, of which it reproduces the characteristic loop between past and future. Importantly, one must note that these two goals, for an image to be both optimistic and credible, are seen as entering in a contradiction. Yet another contradiction arises from the requirement of anticipating a future state early enough, when its features cannot yet be seen clearly, and not waiting until it is too late, when the future is so close to us that it is unchangeable. Both contradictions hint at a necessary balance between the extremes. It is not credible to be too optimistic about the

future, but cognitive paralysis arises when the anticipated future is irreparably catastrophic. It is not credible to announce a prediction too early, but it becomes, not a prediction but a matter of fact, if waited for too long. The methodology of ongoing normative assessment prescribes to *live with* the uncertain future and to follow a certain procedure in continuously evaluating the state of the analyzed system.

The methodology of ongoing normative assessment can also be viewed as a conjunction of inverse prescriptions. This time, instead of an optimistic but credible image of the future, one should wish to obtain at every moment of time an image of the future sufficiently catastrophic to be repulsive and sufficiently credible to trigger the actions that would block its realization. As shown in the discussion of projected time, a closure condition must be met, which takes here the following form: a catastrophe must necessarily be inscribed in the future with some vanishing, but non-zero weight, this being the condition for this catastrophe *not* to occur. The future, on its part, is held as *real*. This means that a human agent is told to *live with* an inscribed catastrophe. Only so will he avoid the occurrence of this catastrophe. Importantly, the vanishing non-zero weight of the catastrophic real future is not the objective probability of the catastrophe and has nothing to do with an assessment of its frequency of occurrence. The catastrophe is altogether inevitable, since it is inscribed in the future: however, if the methodology of ongoing normative assessment is correctly applied, the catastrophe will not occur. A damage that will not occur must be *lived with* and treated *as if* inevitable: this is the aporia of our human condition in times of impending major threats.

To give an example of how ongoing normative assessment is applied in actual cases, we cite the Metropolitan Police commissioner Sir John Stevens, who, speaking about terrorist attacks in London as reflected in his everyday work, said in March 2004, "We do know that we have actually stopped terrorist attacks happening in London but… there is an *inevitablity* that some sort of attack will get through but my job is to make sure that *does not* happen" (Stevens 2004).

Each term in the formulation of the methodology of ongoing normative assessment requires clarification. We start with the word *ongoing*. The assessment that we are speaking about implies systems where the role of the human observer (individual or collective) is the one of observer-participant. As discussed in Section 3.2, the observer-participant does not analyze the system that he interacts with in terms of linear time; instead, he is constantly involved in an interplay of mutual constraints and interrelations between the system being analyzed and himself. The temporality of this relation is the circular temporality of projected time: if viewed from an external, Archimedes' point, influences go both ways, from the system to the observer and from the observer to the system. The observer, who preserves his identity throughout the whole development and whose point of view is 'from the inside', is bound to reason in a closed loop temporality, the only one that takes into account the mutual character of the constraints. Now, if one is to transpose the observer's circular vision back into the linearly developing, occurring

time, he finds that the observer cannot do all his predictive work at one and only one point of occurring time. Circularity of relations within a complex system requires that the observer constantly revise his prediction. To make sure that the loop of interrelations between the system and himself is updated consistently and does not lead to a catastrophic elimination of any major component of either the system in question or of the observer himself, the latter must not stop addressing the question of the future at all times. No fixed-time prediction conserves its validity due to the circularity and self-referentiality of the complex system.

We now address the next term in the formulation of our methodology, *normative* assessment. A serious deficiency of the precautionary principle is that, unable to depart from the normativity proper to the calculus of probabilities, it fails to capture what constitutes the essence of ethical normativity concerning choice in a situation of uncertainty. We argue that judgements are normative but that this normativity, applied to the problem of the future, takes on a special form.

We refer to the concept of 'moral luck' in moral philosophy. Let us first illustrate with an example why probabilistic reasoning does not lead to any satisfactory account of judgement. Imagine that one must reach into an urn containing an indefinite number of balls and pull one out at random. Two thirds of the balls are black and only one third are white. The idea is to bet on the color of the ball before seeing it. Obviously, one should bet on black. And if one pulls out another ball, one should bet on black again. In fact, one should *always* bet on black, even though one foresees that one out of three times on average this will be an incorrect guess. Suppose that a white ball comes out, so that one discovers that the guess was incorrect. Does this *a posteriori* discovery justify a retrospective change of mind about the rationality of the bet that one made? No, of course not; one was right to choose black, even if the next ball to come out happened to be white. Where probabilities are concerned, the information as it becomes available can have no conceivable retroactive impact on one's judgement regarding the rationality of a past decision made in the face of an uncertain or risky future. This is a limitation of probabilistic judgement that has no equivalent in the case of moral judgement.

Take another example. A man spends the evening at a cocktail party. Fully aware that he has drunk more than is wise, he nevertheless decides to drive his car home. It is raining, the road is wet, the light turns red, and he slams on the brakes, but a little too late: after briefly skidding, the car comes to a halt just past the pedestrian crosswalk. Two scenarios are possible: either there was nobody in the crosswalk, and the man has escaped with no more than a retrospective fright. Or else the man ran over and killed a child. The judgement of the law, of course, but above all that of morality, will not be the same in both cases. Here is a variant: the man was sober when he drove his car. He has nothing to reproach himself for. But there is a child whom he runs over and kills, or else there is not. Once more, the unpredictable outcome will have a retroactive impact on the way the man's conduct is judged by others and also by the man himself. Therefore, moral luck

becomes an argument proving that ethics is necessarily a *future ethics*, in Jonas's sense as described earlier, when it comes to judgement about a future event. However, the implementation of that future ethics is impeded in practice by the very inevitability of the uncertainty of the future. This is the ethical aporia we started with.

Is there a way out? Hans Jonas's credo is that there is no ethics without metaphysics. Only a radical change in metaphysics can allow us to escape from the ethical aporia. The major stumbling block of our current, implicit metaphysics of temporality turns out to be our common conception of the *future as unreal*. From the human belief in free will—'we may act otherwise'—is derived the conclusion that the future is not real, in the philosophical sense: 'future contingents', i.e. propositions about actions taken by a free agent in the future, e.g. 'John will pay back his debt tomorrow', are held to have no truth value. They are neither true nor false. If the future is not real, then it is not something that we can have cognizance of. If the future is not real, then it is not something that projects its shadow onto the present. Even when we know that a catastrophe is about to happen, we do not believe it: we do not believe what we know. If the future is not real, there is nothing in it that we should fear, or hope for. From our point of view, the derivation from free will to the unreality of the future is a sheer logical fallacy.

Like the car driver, but on an entirely different scale, human society taken as a collective subject has made a choice in the development of its potential capabilities that brings it under the jurisdiction of moral luck. It may be that its choice will lead to great and irreversible catastrophes; it may be that it will find the means to avert them, to get around them, or to get past them. No one can tell which way it will go. Judgement can only be retrospective. However*, it is possible to anticipate, not the judgement itself, but the fact that it must depend on what will be known once the 'veil of ignorance' covering the future is lifted.* Thus, there is still time to insure that our descendants will never be able to say *'too late!'* — a too late that would mean that they find themselves in a situation where no human life worthy of the name is possible.

Retrospective character of judgement means that, on the one hand, application of the existing norms for judging facts and, on the other hand, evaluation of new facts for updating the existing norms and creating new ones, are two complementary processes. While the first one is present in almost any sphere of human activity, the second process prevails over the first and acquires an all-important role in the anticipation of the future. What is a norm is being revised continuously, and at the same time this ever-changing normativity is applied to new facts. It is for this reason that the methodology of ongoing assessment requires that the assessment be normative and that the norms themselves be addressed in a continuous way.

References

Arendt, H. *The Human Condition*, The University of Chicago Press, 1958.

Broderick, D. *The Spike*, Forge, New York, 2001.

Chang, K. "Smaller Computer Chips Built Using DNA as Template." *New York Times*, November 21, 2003.

Dupuy, J.-P. "Common Knowledge, Common Sense." *Theory and Decision* 27 (1989): 37-62.

_____. "Two Temporalities, Two Rationalities: A New Look at Newcomb's Paradox." in Bourgine, P. & Walliser, B. (eds.), *Economics and Cognitive Science*. New York: Pergamon. 1992, 191-220.

_____. (ed.). *Self-deception and Paradoxes of Rationality.* Stanford: C.S.L.I. Publications, 1998.

_____. "Philosophical Foundations of a New Concept of Equilibrium in the Social Sciences: Projected Equilibrium." *Philosophical Studies* 100 (2000a): 323-345.

_____. *The Mechanization of the Mind*. Princeton: Princeton University Press, 2000b.

Ellsberg, D. "Risk, Ambiguity and the Savage Axioms." *Quarterly Journal of Economics* 75 (1961): 643-669.

Feynman R. "There's Plenty of Room At the Bottom." Talk given on at the annual meeting of the American Physical Society at the California Institute of Technology, 1959.

Fleming, D. "The Economics of Taking Care: An Evaluation of the Precautionary Principle." in Freestone, D. & Hey, E. (eds.), *The Precautionary Principle and International Law*. La Haye: Kluwer Law International, 1996.

Jonas, H. *The Imperative of Responsibility. In Search of an Ethics for the Technological Age.* Chicago: University of Chicago Press, 1985.

de Jouvenel, B. *L'art de la conjecture*, Editions du Rocher, Monaco, 1964.

Kourilsky, Ph. & Viney, G. *Le Principe de précaution.* Report to the Prime Minister, Paris, Éditions Odile Jacob, 2000.

Merke, R. http://www.zyvex.com/nano, 2003.

Meyerson, E. *De l'explication dans les sciences* Paris, 1927.

Nordmann, A. (rapp.) *Converging Technologies—Shaping the Future of European Societies*, European Commission report, 2004.

von Neumann, J. "The General and Logical Theory of Automata." Talk given at the Hixon Symposium at the California Institute of Technology, 1948.

Rip, A., Misa, Th. J., & Schot, J. W. (eds.). *Managing Technology in Society. The Approach of Constructive Technology Assessment*. London: Pinter Publishers, 1995.

Roco, M. & Bainbridge, W. (eds.). *Converging Technologies for Improving Human Performances*, National Science Foundation report, 2002.

Stevens, J. http://news.bbc.co.uk/2/hi/uk_news/politics/3515312.stm, 2004.